



eResearch Collaboration Projects – supporting CSIRO's digital science and research

CSIRO Scientific Computing

John Zic, Justin Baker | February 2020





Australia's National Science Agency

5500

talented staff

\$1billion+
budget

Working
with over
2800+
industry
partners

57

sites across
Australia

Top 1%
of global
research
agencies

Each year
CSIRO
contributes
\$4.5 billion+ to
the economy



CSIRO's Challenges & Digital Transformation Program

Key components

CSIRO Missions

The Managed Data Ecosystem

The Digital Academy

Future Science and Technology
(was Decadal Science Plan)

Contribution from

CSIRO Scientific Computing Services

Source: CSIRO MDE Project Team (2019).





CSIRO Missions

Major collaborative scientific and research programs addressing six major challenges facing Australia.



Source: CSIRO MDE Project Team (2019).



Delivering outcomes for Missions

Data and Datasets Management

- Uniform, standardised and operationally supported

Supporting Computational and Data Science

- Expertise partnerships
- On-going learning

Strategic, long-term thinking

**Managed
Data Ecosystem**

**Digital
Academy**

**Future Science
and Technology**



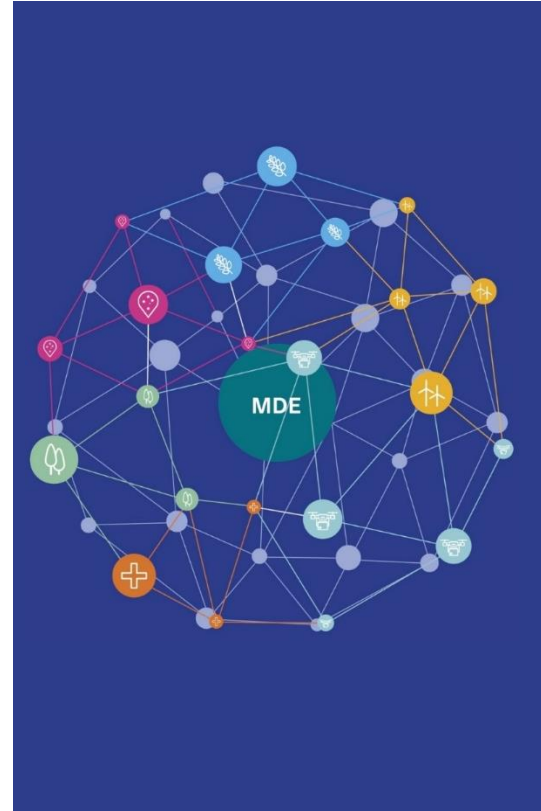
Managed Data Ecosystem (MDE)

CSIRO's scientific and research data and collections are *highly valued*

MDE will provide a uniform set of tools, platforms, policies and governance enabling CSIRO and partners to:

- Share and collaborate on datasets;
- Maintain and understand data provenance and ownership;
- Carry out data mining and analysis of data in a consistent and uniform manner.

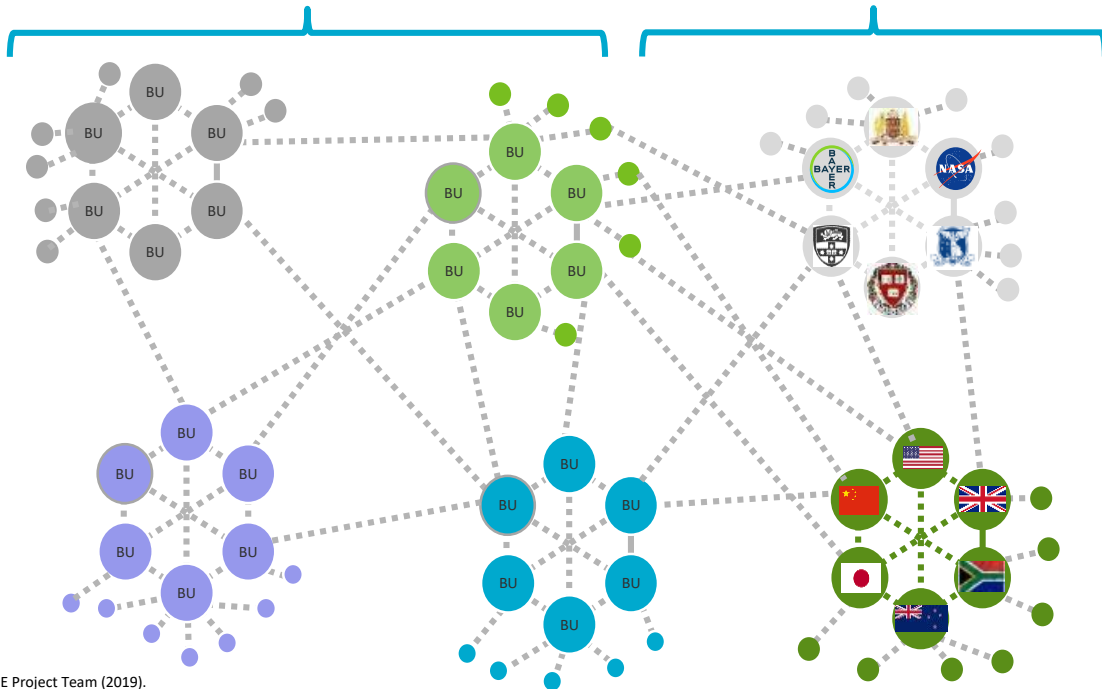
Source: CSIRO MDE Project Team (219).



A Collaborative Future

CSIRO

National and International Research
Community



Source: CSIRO MDE Project Team (2019).

The MDE brings a number of benefits



VALUE

Optimised value for enhanced impact



ACCESS

Clear data sharing standards enabling safe, open access



USE

Complete data provenance



STORAGE

Fit-for-purpose & interconnected solutions



UNDERSTANDING

Comprehensive understanding of data assets



COMPLIANCE

Systematic, fully assured governance



ANALYSIS

Curated, best-of-breed tool kits



Digital Academy

Define and Design

Computational and data science expertise and methodologies

Attract and Retain

A workforce with the computational and data science skills

Learn and Grow

Develop new computational and data science capabilities in our nation's scientists.

Connect and Collaborate

Facilitate digital workforce connectivity, accessibility and collaboration



Scientific Computing



Support program within CSIRO
Total ~90 staff

- Scientific Computing Platforms
- Scientific Computing Services
 - eResearch Collab Projects
 - Pan-deployments
 - BAU



SCIENTIFIC COMPUTING SERVICES

CSIRO Researchers/CSIRO Research Projects

Digital Science Consultancy

Planning & Provisioning;
Analytics;
Science domain computing and data expertise:

Bioinformatics; Astronomy;
Chemistry; Earth Sciences;
Synchrotron Science;

Concierge

Embedding Computational Science Expertise

Driven by requirements from research projects and strategic initiatives

Data Analytics and Visualisation
Research Software Engineering
Technical Solutions
Data Pipelines and Modelling

Capability Development

Digital Learning

Python; Using CSIRO HPC;
ML, AI and Data Analytics;
Software Engineering

Outreach

Conferences - C3DIS+RDA 2020
Student and Internships
Technical Leaders
Drop-in expert sessions

Scientific Software Licensing
License management

Service delivery and partnerships

Software services

Partnership advisor and liaison

Research Data

High Performance & Scalable
Compute Platforms

Partner Facilities



Scientific Computing Services - Expertise

Data analytics & visualisation	Domain specific algorithms & software
Science workflows data and compute centric	Science data management
High performance scientific	Solution Design
Simulation & modelling	Specialised full stack web development
Software Engineering	

Total of 42 staff in 4 Teams

eResearch Collaboration Projects

Delivering Scientific Computing expertise for CSIRO

Research
Software
Engineering



Data Analytics
and
Visualisation



Technical
Solutions



Data Pipelines
and Modelling



- Competitive process – RFP
- Researchers submit ~ 60-80 projects proposals per annum
- ~ 40 projects accepted last cycle
- Projects
 - run for a 6 month cycle
 - 0.2FTE allocated per project and can be applied in various ways

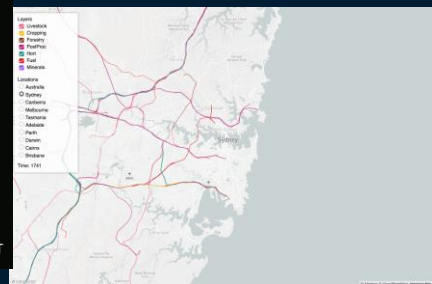
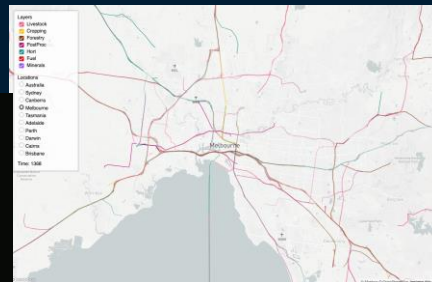
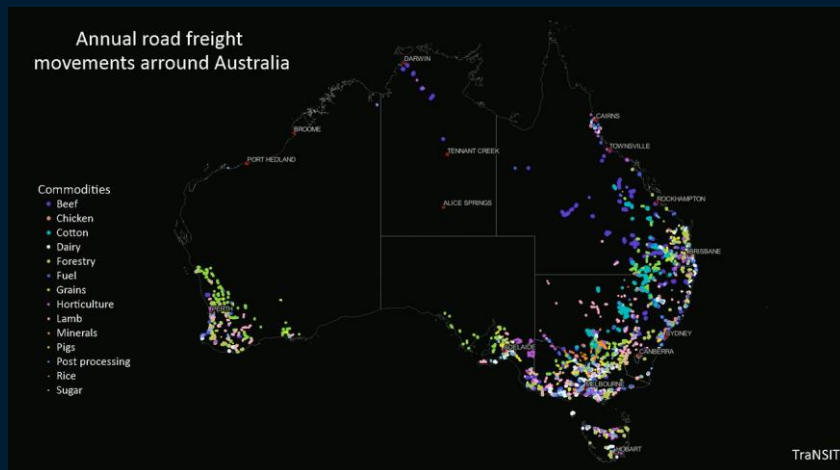


Example Project – Freight transit across Australia

CSIRO developed transport logistics tool mapping and visualising millions of vehicle trips between production and domestic and export markets.

Data Analytics and Visualisation team: Louise Ord (IM&T, South Eveleigh)

TraNSIT team: Adam McKeown (L&W, Cairns) and Artio Bondarenco (L&W, Dutton Park)

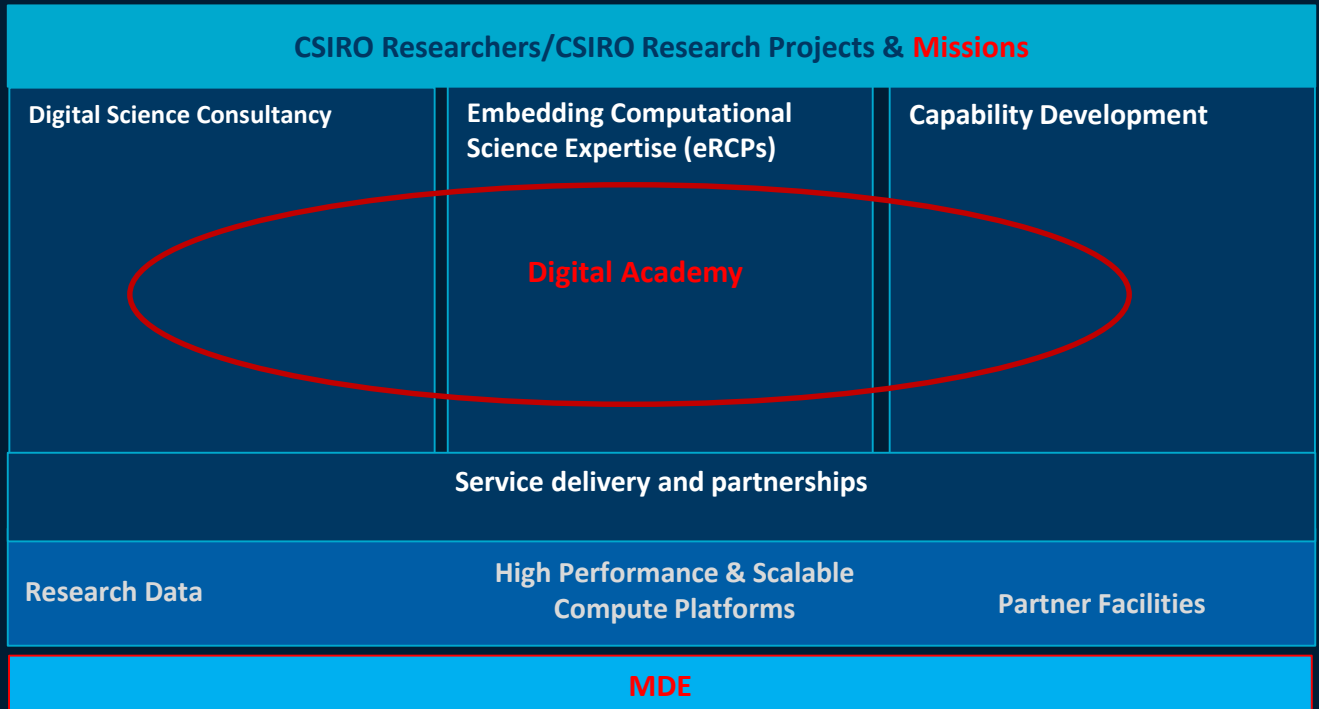


Scientific Computing Services - Future Plans

Current	Future
BAU	BAU
eRCPs	Some reduction in eRCPs
Pan-deployments	Reduction/Discontinuation
Client training via Digital Academy	Broader range of Academy contributions
-	Targeted MDE pilots, MDE projects
-	Missions through Business Unit engagement



Scientific Computing Services - Future Plans





Thank you

Justin Baker

Team Leader, Data
Analytics and
Visualisation

T +61 3 9545 2012
E justin.baker@csiro.au

Dr John Zic

Executive Manager,
Scientific Computing
Services

T +61 2 9490 5682
E john.zic@csiro.au

Challenges

Resourcing:

- Allocation complexity:
staff ↔ projects
- Matching skills ↔ project requirements

Variable recognition of contributions to project outcomes

Occasional project failures



eResearch Collaboration Projects - Future Plans

MDE

- Redirect specialist expertise to pilots
- Contribute to ongoing operational rollout

Digital Academy

- Develop/adapt training content
 - Eg HPC software carpentry (customise)
- Deliver training material to staff
- Understand/identify skills for new talent

Missions

- Continue to provide eResearch support as needed

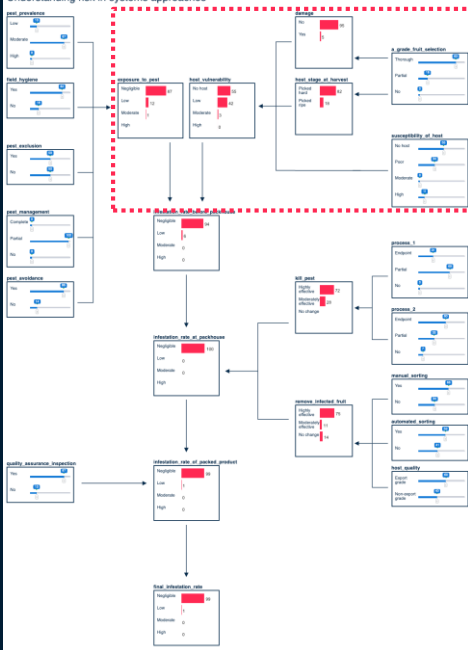


Example Project 2

Bayesian Network for understanding systems risk

Bayesian Networks

Understanding risk in systems approaches

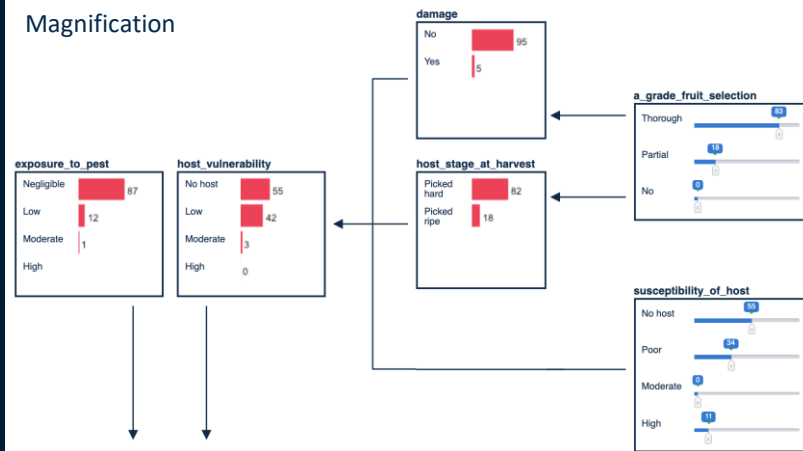


Shiny dashboard enables clients to interact with expert knowledge-driven Bayesian network – allows model parameter selection and and scenario analysis without the need to interact with the underlying model.

Data Analytics and Visualisation team: Louise Ord (IM&T, South Eveleigh)

Modelling team: Justine Murray (H&B, Dutton Park) and Bill Venables (Data61, Dutton Park)

Magnification





Digital science underpins everything we do

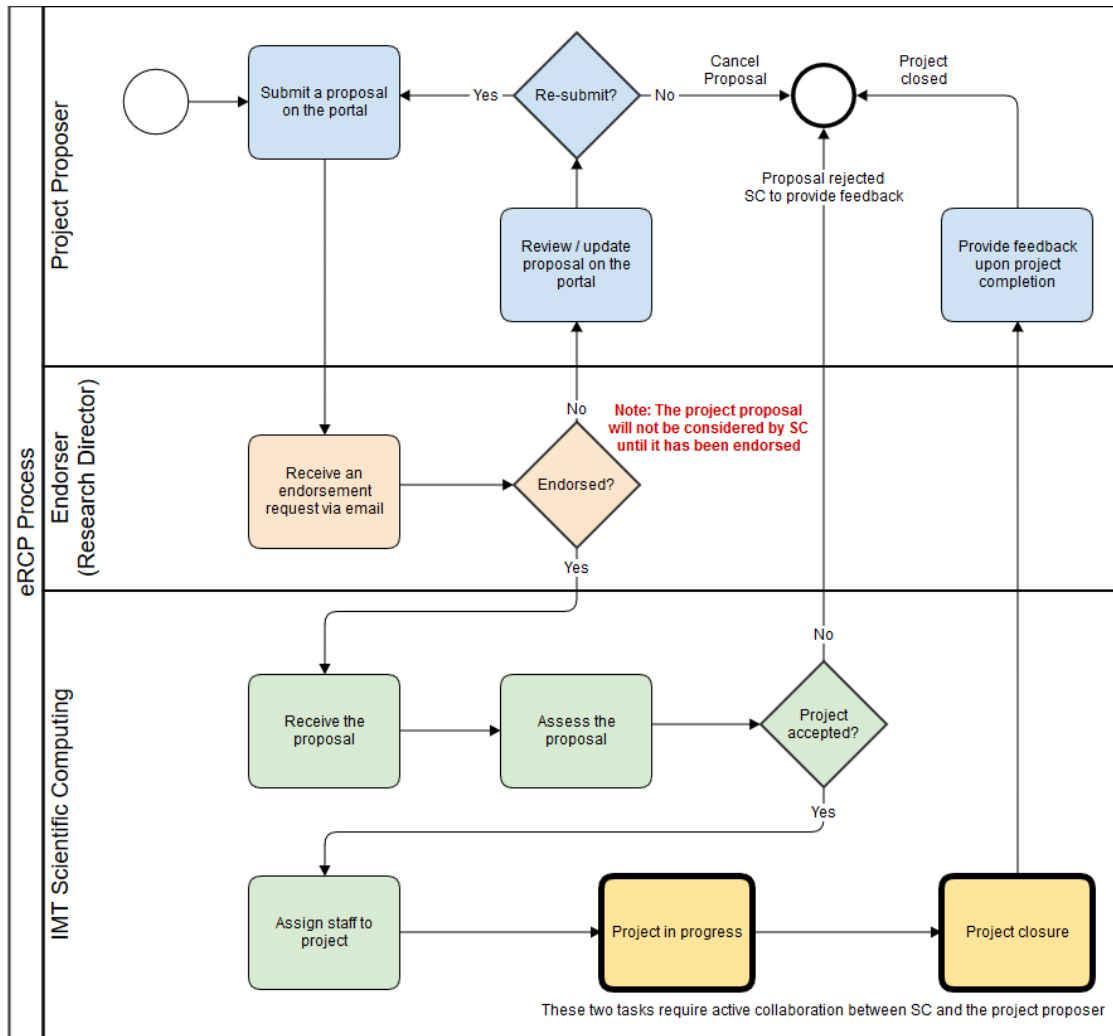
To successfully advance the national research agenda, our team requires the most cutting-edge tools and approaches available.

Digital science is all about:

- **Computer models** – digital twins of – simulations
- **Algorithms** that learn and can direct actions – autonomous robots
- **Big data** that can be mined for trends, to find things in the data or to predict outcomes

Source: CSIRO MDE Project Team (2019).







Project Submission Portal



eRCP Proposals

Previous 1 2 Next					☐ Show Proposals Pending Submission		Search Proposals...
Activity / Project Title	Proposer	Jira #	Round	Status			
 Mquest - improving a visual quality control tool for ocean temperature data	Cowley, Rebecca (O&A, Hobart)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		
 Tools for defining chemical crystallisation space	Newman, Janet (Manufacturing, Parkville)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		
 AI for image analysis	Newman, Janet (Manufacturing, Parkville)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		
 Facilitating access to Sentinel 1/2 satellite data	Poulsen, Brett (Energy, Pullenvale)	ERRFP-652	2018 July - Dec	Application Received	VIEW PROPOSAL		
 Real Time Soft Matter Modeling for Autonomous Design	Hockings, Nick (Data61, Pullenvale)	ERRFP-651	2018 July - Dec	Application Received	VIEW PROPOSAL		
 Development of image analysis algorithms for automated tag reading	Wang, Dadong (Data61, Marsfield)	ERRFP-650	2018 July - Dec	Application Received	VIEW PROPOSAL		
 Biomass Quality Database	Roberts, Daniel (Energy, Pullenvale)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		
 Immersive Marine Sensor Data Analytics	Engelke, Ulrich (Data61, Sandy Bay)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		
 Processing electronic health records using deep learning	Karimi, Sarvnaz (Data61, Marsfield)		2018 July - Dec	Awaiting Endorser	VIEW PROPOSAL		